Subject: Re: Correct weight for a sub-sample
Posted by Reduced-For(u)m on Thu, 30 May 2013 01:42:22 GMT
View Forum Message <> Reply to Message

I think it totally depends on what sub-population you are looking at. If you are looking at the bottom quintile of the asset index, I'm not sure that that is the kind of sub-population you want to weight in terms of probability of sampling. They are by definition (I think) the lowest 20% of scores from a principal component analysis. I don't think those are weighted in any probability sense when computed (meaning, if you tab out the quintiles, I think you actually just see 20% in each bin, unweighted - though in some of the newer surveys I think they do this differently between rural and urban households, but that is adding on another layer of complexity).

So I'm just not sure that there are "Nationally Representative Weights for the Bottom Quntile of Household Asset Index". What would "nationally representative" mean in that context?

As for just the STATA question though, one thing you could do is something like this, which would preserve the relative probabilities implied in the DHS weights for your sample.

```
***begin
gen preweight = v005/100000
keep if assetquintile==5
egen weightsum = total(weight)
gen newweight = preweight/weightsum
```

*now you have weights that add up 1 for the group you wanted, proportional to their original weights. I'm not sure the interpretation is just what you wanted, but I'm not sure there is a perfect interpretation of what you want either.

*now, for your regressions, you can either just set this as the new weight in the same svyset manner
*Another option would be to directly specify the estimating procedure using [pweight=weight] and an appropriate clustering level.

***End

Interestingly, I think this answers your other question too, about cross-country stuff. Since the DHS weights sum to N (sample size) you need to re-normalize them so they all add up to 1 for each country...or, depending on what you are trying to do in that cross-country thing, maybe re-scale them again so that they sum up to Population. Depends on the parameter you are trying to estimate and the assumptions you're willing to make. We can pick this up in the other thread you commented on if you'd like.